

**Test  
du 29/11/2021**  
Durée : 15 minutes

Nom : .....

Prénom : .....

Parcours :  DS  IMSA  MASS POP

Des points négatifs pourront être affectés à des mauvaises réponses. Cochez les réponses avec un stylo noir ou bleu foncé, et utiliser du blanc pour corriger vos erreurs.

**Question 1** Pour choisir/sélectionner des covariables, on peut utiliser

- lasso       ridge       l'analyse de la variance

**Question 2** Un effet  $\beta_j$  en facteur d'une covariable binaire  $X_j$  s'interprète comme une différence de moyennes sur des sous-populations

- Vrai       Faux

**Question 3** En cas de sur-apprentissage,

- $\hat{\sigma}^2$  sous-estime  $\sigma^2$ , la variance de l'erreur
- on utilise la validation croisée pour supprimer ce phénomène
- l'estimation  $\hat{\boldsymbol{\beta}}$  est une très bonne estimation de  $\boldsymbol{\beta}$
- aucune des trois autres options

**Question 4** Dans une régression, on doit remplacer une covariable catégorielle  $X_j$  par autant de variables binaires que de modalités de  $X_j$

- Vrai       Faux

**Question 5** Dans le modèle de régression linéaire ajusté par moindres carrés, centrer-reduire les covariables change les prédictions sur la variable réponse

- Vrai       Faux

**Question 6** Pour utiliser l'hypothèse gaussienne, il est utile de vérifier que

- $Y$  est gaussienne
- $R^2$  n'est pas trop proche de 0
- les covariables  $X_1, \dots, X_p$  sont gaussiennes
- aucune des trois autres options

**Question 7** On considère une variable  $Y$  d'espérance  $\theta$  et de variance  $\sigma^2$ . On estime  $\theta$  par la moyenne de l'échantillon  $\hat{\theta}$  de l'échantillon. Soit  $Y_{\text{new}}$  indépendant de l'échantillon. On a  $E((E(Y_{\text{new}}) - \hat{\theta})^2) = \sigma^2/n$ .

- Vrai       Faux

**Question 8** Avec une méthode rétrograde de sélection de variables, le critère AIC conserve plus ou autant de covariables que le critère BIC

- Vrai       Faux